УДК 004.853
https://doi.org/10.23947/2541-9129-2019-3-17-22

ПРОГНОЗИРОВАНИЕ ПЛОЩАДИ ГОРЕНИЯ ЛЕСНОГО ПОЖАРА С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ

Филиппенко В. А., Зотов А. В.
Донской государственный технический университет, Ростов-на-Дону, Российская Федерация
philippenko@inbox.ru,
azotzot@gmail.com

Целью настоящей работы является создание и обучение искусственной нейронной сети на основе набора данных, содержащих различные климатические параметры и будущую площадь пожара в качестве выходного прогнозируемого параметра. Такой набор данных является, как правило, доступным для исследования и изучения. Перед обучением модели нейронной сети набор данных разделяют на две выборки — выборка для обучения, которая составляет около 90 % от набора, и выборка для тестирования обученной модели. В постановке задачи авторы выбирают и анализируют известные данные о пожарах в парке Монтезиньо (*Montesinho*), сравнивают модели, обученные на этих данных с нормализацией и без нее. В качестве результата приведены два примера графиков изменения абсолютной ошибки площадей пожара, прогнозируемых с помощью созданной и обученной модели.

Ключевые слова: площадь горения, машинное обучение, модель, нейронные сети, *Keras*, прогнозирование, лесной пожар.

UDC 004.853
https://doi.org/10.23947/2541-9129-2019-3-17-22

FORECASTING FOREST FIRE BURNING AREA USING MACHINE TRAINING

Filippenko V. A., Zotov A. V.
Don State Technical University, Rostov-on-Don, Russian Federation
philippenko@inbox.ru,
azotzot@gmail.com

The objective of this article is to create and train an artificial neural network based on a data set containing various climatic parameters and future fire area as an output parameter that the authors intend to predict. Such a "set" of data is usually available for research and study. Before training the neural network model, the data set is divided into two samples: a sample for training, which is about 90% of the set; and a sample for testing the trained model. In setting the task, the authors select and analyze the known data on the fires that occurred in Montesinho Park, compare the models trained on these data with and without normalization. As a result, two examples are given of a qualitative demonstration of graphs of absolute error changes of fire areas, which are projected using the created and trained model.

Keywords: burning area, machine training, model, neural networks, *Keras*, forecasting, forest fire.

Введение. Лесным пожаром называют стихийное и неуправляемое распространение огня по лесным площадям. Согласно данным Федерального агентства лесного хозяйства за неделю с 3 по 9 июня 2019 года в 45 регионах России лесопожарные силы и привлеченные лица ликвидировали 354 лесных пожара на площади 5783,2 га, в том числе за выходные дни 8–9 июня потушено 98 пожаров на площади 1790,05 га. Из-за дыма при возгораниях ежегодно умирают около 300 тысяч человек. В результате сгорании биомассы образуется аэрозольно-газовая смесь, представляющая эколого-токсикологический риск для человека.

Сотрудники пожарной охраны должны быть обеспечены максимально эффективным пожарно-техническим снаряжением и техникой для ликвидации стихийных явлений. Но часто этого недостаточно для эффективной борьбы с таким опасным явлением. Стратегическое планирование

и распределение ресурсов, например предоставление достаточного количества пожарных самолётов или наземных бригад, могут существенно повысить шансы при борьбе с пожаром. Но для этого нужно рассчитать количество ресурсов, что может занять довольно много времени.

Одним из способов решения данной проблемы может быть использование нейронных сетей. В представленной работе для обучения и тестирования модели нейронной сети авторы использовали данные о произошедших пожарах в парке Монтезиньо (*Montesinho*), расположенном в Португалии. Этот набор данных общедоступен для исследования и работы с ним [1]. Авторы используют нейросетевую библиотеку *Keras* [2], написанную на языке программирования *Python* [3, 4].

Подготовка данных. Данные возгорания парка Монтезиньо были выбраны в качестве обучающего материала для модели нейронной сети в связи с тем, что комплексный показатель пожарной опасности В. Г. Нестерова, используемый в РФ, содержит меньше параметров, а это может быть причиной более низких результатов при обучении модели. В используемом авторами наборе параметров, кроме традиционных, содержатся параметры: влагосодержание лесной подстилки и почвы, характеристики пламени, антропогенный фактор и грозовая активность. Ниже приведены дополнительные параметры рейтинговой системы лесной пожарной опасности, которые использовались при формировании данного набора [5]:

- вероятность возгорания (*Fine Fuel Moisture Code, FFMC*);
- норма влажности угля (*Duff Moisture Code, DMC*);
- норма засухи (*Drought Code, DC*);
- индекс первоначального распространения системы (*Initial Spread Index, ISI*).

Все метеорологические данные для расчёта упомянутых выше компонентов можно запросить у ближайшей метеорологической службы. Так как данный набор данных содержит довольно много климатических параметров, с помощью созданной и обученной модели можно будет предсказывать будущую площадь пожара не только для парка Монтезиньо, но и на любой другой схожей с ним территории.

Полные данные, содержащиеся в наборе:

- *X* — пространственная координата оси *X* на карте парка Монтезиньо: от 1 до 9;
- *Y* — пространственная координата оси *Y* на карте парка Монтезиньо: от 2 до 9;
- «месяц» — месяц года: от января до декабря;
- «день» — день недели: с понедельника по воскресенье
- *FFMC* — индекс легкости воспламенения топлива из системы *FWI* в интервале значений 18,7–96,2;
- *DMC* — индекс нормы влажности угля из системы *FWI* в интервале значений 1,1–291,3;
- *DC* — индекс нормы засухи от системы *FWI* в интервале значений 7,9–860,6;
- *ISI* — индекс первоначального распространения из системы *FWI* в интервале значений от нуля до 56,1;
- «Температура» — температура в интервале значений 2,2–33,3°C;
- относительная влажность от 15,0 до 100 %;
- «Ветер» — скорость ветра от 0,4 до 9,4 км/ч;
- дождь снаружи от 0,0 до 6,4 мм/м^2;
- «Площадь» — сожженная площадь леса от 0,00 до 1090,84 га.

Все параметры в наборе изменяются в разных диапазонах. Для того, чтобы повысить точность прогнозирования модели, нужно провести нормализацию данных. Один из способов нормализировать данные — это вычесть из каждого параметра его среднее значение и разделить на

стандартное отклонение. После этих действий среднее значение будет являться нулём, а дисперсия — единицей. В этом случае данные в каждом столбце будут изменяться от –1 до +1, но при таком способе нормализации в некоторых столбцах могут появиться отрицательные значения, что не может быть для некоторых параметров. Для решения этой проблемы можно использовать процедуру *MinMaxScaler().fit_transform()* [6], которая преобразует все данные в диапазон 0 … +1. Такая модель обучается с помощью «обучения с учителем». В этом случае данные делятся на две части — данные для обучения и верные ответы для этих данных. Данные для обучения необходимы для обучения модели, а ответы — для перерасчёта весов на рёбрах графа нейросети при несовпадении предсказанного значения и действительного. Перед обучением случайным образом разделим эти данные на обучающую выборку и выборку тестовую. Обучающая выборка является частью набора данных, используется для обучения модели. Она будет составлять около 90 % набора. Тестовая модель составляет 10 % набора данных и служит для проверки работоспособности модели. Тестовые данные не будут участвовать в обучении модели, они служат лишь для проверки работоспособности. В последующем такое моделирование можно связать с классическим построением моделей и расчетом показателей техносферной безопасности [7,8].

Создание модели. В процессе тестирования различных моделей для рассматриваемого набора данных лучше всего зарекомендовала себя модель, содержащая 6 слоёв: входной слой с 24 нейронами, 4 скрытых слоя, содержащих 48, 96, 48, 24 слоя, и выходной нейрон.

Использовались следующие функции активации:

• линейная — на первом, втором и пятом слоях;

• сигмоида — на втором и третьем слоях, она позволяет усиливать слабые сигналы, не насыщаясь сильными;

• *selu* — на выходном нейроне, повышает показатель сходимости нейронной сети.

При компиляции модели в качестве оптимизатора типа градиентного спуска использовалась «*adadelta*». *Adadelta обновляет меньшие веса, которые слишком часто обновляются, но, в отличии от Adagrad, вместо полной суммы обновлений будет использоваться усреднённый по истории квадрат градиента.*

В качестве функции ошибки, которая будет использоваться оптимизатором в алгоритме обратного распространения ошибки, выбираем среднеквадратичную ошибку, в качестве метрики — «mae», среднюю абсолютную ошибку.

Обучение. *На 500-эпохе обучения средняя абсолютная ошибка равняется 4,6, значит модель при предсказаниях будет ошибаться в целом на 4,6 гектара, что авторы считают удовлетворительным. Кривая изменения ошибки представлена на рис. 1.*

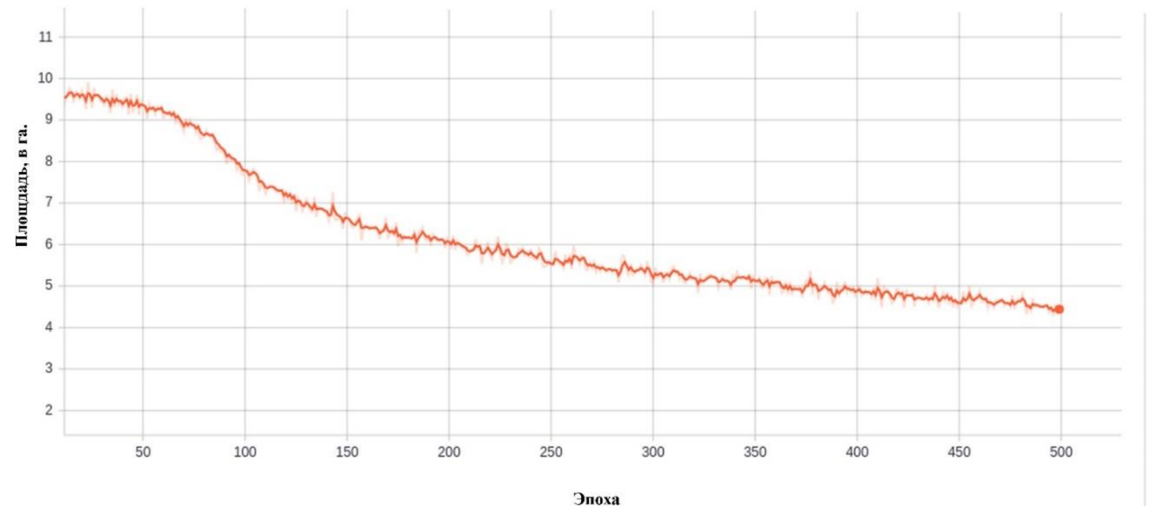


Рис. 1. Кривая изменения абсолютной ошибки при обучении на данных с нормализацией

На рис. 2 изображен график изменения абсолютной ошибки при ненормализованных данных, который доказывает, что нормализованные данные показывают себя лучше, чем исходные.

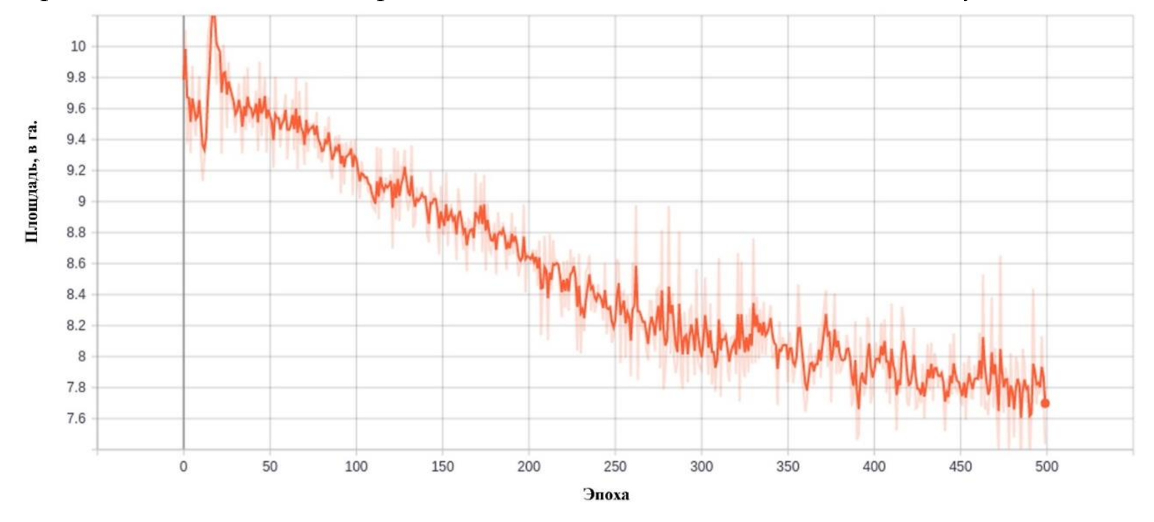


Рис. 2. Кривая изменения абсолютной ошибки при обучении на данных без нормализации

Прогнозирование. *На рис. 3 представлен график, иллюстрирующий результат работы нейронной сети. На графике оранжевая линия — действительная площадь горения, синяя линия — предсказанная площадь. Как видно на рисунке, динамика изменения кривых для каждой записи практически идентична, что показывает хорошую работу обученной модели при прогнозировании.*

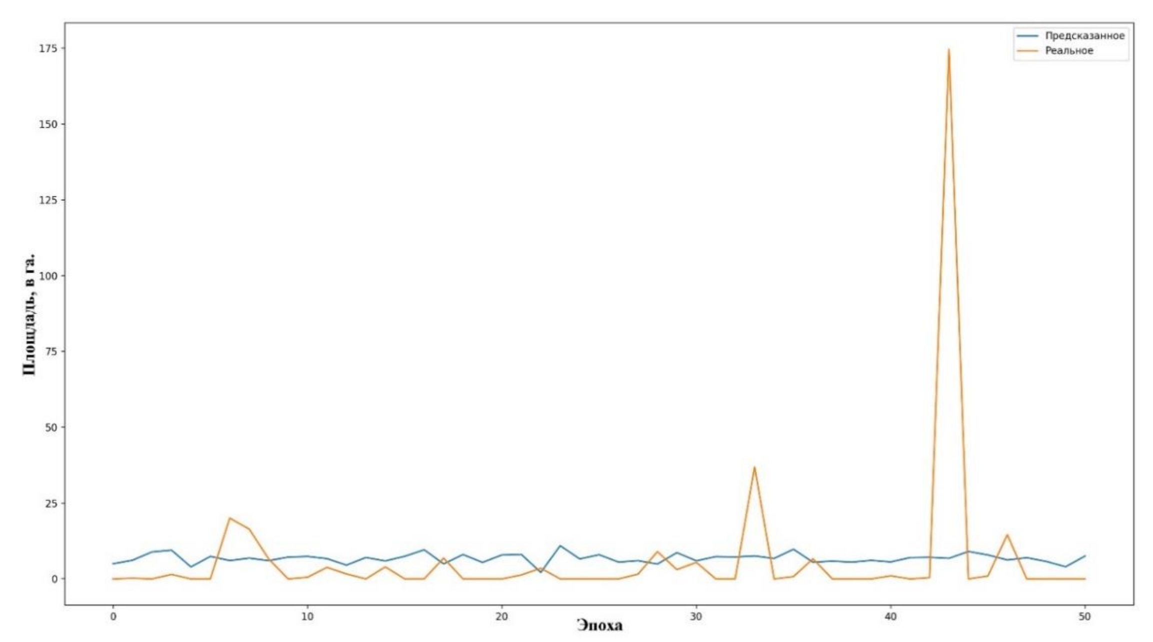


Рис. 3. Прогнозирование с помощью обученной модели на данных с нормализацией

На рис. 4 продемонстрирован график прогнозирования другой модели, которая обучалась с помощью ненормализованных данных. Очевидно, что первая модель более эффективна.

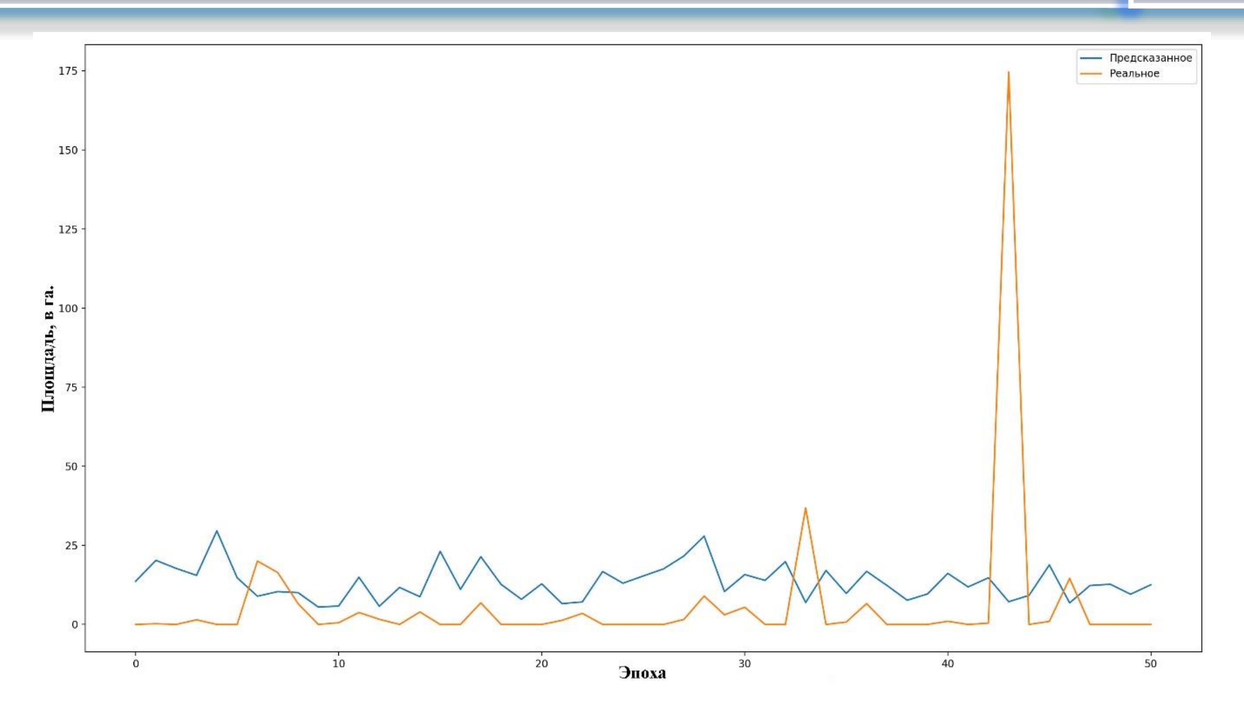


Рис. 4. Прогнозирование с помощью обученной модели на данных без нормализации

Заключение. В данной работе создана и обучена модель искусственной нейронной сети на наборе данных, содержащих различные климатические параметры и будущую площадь пожара в гектарах. Эта площадь является выходным параметром, который авторы собираются прогнозировать. Как правило, такой набор данных доступен для исследования и изучения. Перед обучением модели нейронной сети набор данных разделяли на две выборки: выборка для обучения, которая составляет около 90 % от набора, и выборка для тестирования обученной модели. В постановке задачи авторы выбирают и анализируют известные данные о произошедших пожарах в парке Монтезиньо (*Montesinho*), сравнивают модели, обученные на этих данных, с нормализацией и без нее. В качестве результата приведены два примера демонстрации графиков изменения абсолютной ошибки площадей пожара, прогнозируемых с помощью созданной и обученной модели.

Библиографический список.

1. Cortez, P. A Data Mining Approach to Predict Forest Fires using Meteorological Data / P. Cortez, A. Morais [Электронный ресурс] // Dep. Information Systems/Algoritmi R&D Centre University of Minho. — Режим доступа: http://www3.dsi.uminho.pt/pcortez/fires.pdf (дата обращения: 10.06.2019).

2. Библиотеки для глубокого обучения: Keras [Электронный ресурс] // Open Data Science. — Режим доступа : https://habr.com /ru/company /ods/blog /325432/ (дата обращения: 10.06.2019).

3. Лутц, М. Изучаем Python / М. Лутц. — Санкт-Петербург : Символ-Плюс, 2011. — 1280 с.

4. Мюллер, А. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными / А. Мюллер, С. Гвидо. — Санкт-Петербург : ООО «Альфа-книга», 2017. — 480 с.

5. Губенко, И. М. Сравнительный анализ методов расчета индексов пожарной опасности / И. М. Губенко, К. Г. Рубинштейн [Электронный ресурс] // Гидрометеорологический научно-исследовательский центр Российской Федерации. — Режим доступа : http:// method.meteorf.ru /publ/tr/tr347 /gubenko.pdf (дата обращения: 10.06.19).

6. Кин, Б. А Масштабирование функций с помощью scikit-learn [Электронный ресурс] / Бен Алекс Кин. — Режим доступа : http:// benalexkeen.com/ feature-scaling-with-scikit-learn/ (дата обращения: 10.06.2019).

7. Есипов, Ю. В Логическое и параметрическое моделирование предпосылок и установление меры определенности реализации происшествия в системе / Ю. В. Есипов, М. С. Джиляджи, А. И. Черемисин // Безопасность в техносфере — 2017. — Т. 6, № 2. — С. 3–11.

8. Есипов, Ю. В. Модели и показатели техносферной безопасности / Ю. В. Есипов, Ю. С. Мишенькина, А. И. Черемисин. — Москва : ИНФРА, 2018. — 154 с.

Об авторах:

Филиппенко Виктор Александрович,
студент *Донского государственного технического университета, (РФ, г. Ростов-на-Дону, пл. Гагарина,1),*
philippenko@inbox.ru

Зотов Алексей Вячеславович,
студент *Донского государственного технического университета, (РФ, г. Ростов-на-Дону, пл. Гагарина,1),*
azotzot@gmail.com